

--	--	--	--	--	--	--	--	--	--

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2022/2023 (TERM ID 2210)

BEC2034 – BASIC ECONOMETRICS

(All Sections/ Groups)

13 FEBRUARY 2023
09:00 A.M. – 11:00 A.M.
(2 Hours)

INSTRUCTIONS TO STUDENTS

1. This Question paper consists of **EIGHT** pages with **TWO** sections, excluding formula sheet and statistical tables.
Section A – Twenty (25) Multiple Choice Questions
Section B – Two (2) Structured Questions
2. Attempt **ALL** questions in both **Sections A and B**. The distribution of the marks for each question is given.
3. Please write all your answers in the Answer Booklet provided.

SECTION A (MULTIPLE CHOICE QUESTIONS) (50 marks)

1. Heteroskedasticity means that _____.
 - A) homogeneity cannot be assumed automatically for the model
 - B) the variance of the error term is not constant
 - C) the observed units have different preferences
 - D) agents are not all rational
2. Under the least squares assumptions (zero conditional mean for the error term, X_i and Y_i being i.i.d., and X_i and u_i having finite fourth moments), the OLS estimator for the slope and intercept _____.
 - A) has an exact normal distribution for $n > 15$
 - B) is BLUE
 - C) has a normal distribution even in small samples
 - D) is unbiased
3. One of the following steps is not required as a step to test for the null hypothesis _____.
 - A) compute the standard error of $\hat{\beta}_1$
 - B) test for the errors to be normally distributed
 - C) compute the t-statistic
 - D) compute the p-value
4. A binary variable is often called a _____.
 - A) dummy variable
 - B) dependent variable
 - C) residual
 - D) power of a test
5. The homoskedastic normal regression assumptions are all of the following with the exception _____.
 - A) the errors are homoskedastic
 - B) the errors are normally distributed
 - C) there are no outliers
 - D) there are at least 10 observations
6. When there are omitted variables in the regression, which are determinants of the dependent variable, then _____.
 - A) you cannot measure the effect of the omitted variable, but the estimator of your included variable(s) is (are) unaffected
 - B) this has no effect on the estimator of your included variable because the other variable is not included
 - C) this will always bias the OLS estimator of the included variable
 - D) the OLS estimator is biased if the omitted variable is correlated with the included variable

Continued...

7. Imagine you regressed earnings of individuals on a constant, a binary variable ("*Male*") which takes on the value 1 for males and is 0 otherwise, and another binary variable ("*Female*") which takes on the value 1 for females and is 0 otherwise. Because females typically earn less than males, you would expect _____.
A) the coefficient for Male to have a positive sign, and for Female a negative sign
B) both coefficients to be the same distance from the constant, one above and the other below
C) none of the OLS estimators to exist because there is perfect multicollinearity
D) this to yield a difference in means statistic
8. In the multiple regression model, the least squares estimator is derived by _____.
A) minimising the sum of squared prediction mistakes
B) setting the sum of squared errors equal to zero
C) minimising the absolute difference of the residuals
D) forcing the smallest distance between the actual and fitted values
9. The main advantage of using multiple regression analysis over differences in means testing is that the regression technique _____.
A) allows you to calculate p-values for the significance of your results
B) provides you with a measure of your goodness of fit
C) gives you quantitative estimates of a unit change in X
D) assumes that the error terms are generated from a normal distribution
10. If the absolute value of your calculated *t*-statistic exceeds the critical value from the standard normal distribution you can _____.
A) safely assume that your regression results are significant
B) reject the null hypothesis
C) reject the assumption that the error terms are homoskedastic
D) conclude that most of the actual values are very close to the regression line
11. If you reject a joint null hypothesis using the *F*-test in a multiple hypothesis setting, then _____.
A) a series of *t*-tests may or may not give you the same conclusion
B) the regression is always significant
C) all of the hypotheses are always simultaneously rejected
D) the *F*-statistic must be negative
12. Panel data estimation can sometimes be used _____.
A) to avoid the problems associated with mis-specified functional forms
B) in case the sum of residuals is not zero
C) in the case of omitted variable bias when data on the omitted variable is not available
D) to counter sample selection bias

Continued...

13. When your multiple regression function includes a single omitted variable regressor, then _____.
A) use a two-sided alternative hypothesis to check the influence of all included variables
B) the estimator for your included regressors will be biased if at least one of the included variables is correlated with the omitted variable
C) the estimator for your included regressors will always be biased
D) lower the critical value to 1.645 from 1.96 in a two-sided alternative hypothesis to test the significance of the coefficients of the included variables
14. Misspecification of functional form of the regression function _____.
A) is overcome by adding the squares of all explanatory variables
B) is more serious in the case of homoskedasticity-only standard error
C) results in a type of omitted variable bias
D) requires alternative estimation methods such as maximum likelihood
15. Time series variables fail to be stationary when _____.
A) the economy experiences severe fluctuations
B) the population regression has breaks
C) there is strong seasonal variation in the data
D) there are no trends
16. Departures from stationarity _____.
A) jeopardize forecasts and inference based on time series regression
B) occur often in cross-sectional data
C) can be made to have less severe consequences by using log-log specifications
D) cannot be fixed
17. The notation for panel data is $(X_{it}, Y_{it}), i = 1, \dots, n$ and $t = 1, \dots, T$ because _____.
A) we take into account that the entities included in the panel change over time and are replaced by others
B) the X's represent the observed effects and the Y the omitted fixed effects
C) there are n entities and T time periods
D) n has to be larger than T for the OLS estimator to exist
18. The Augmented Dickey Fuller (ADF) t -statistic _____.
A) has a normal distribution in large samples
B) has the identical distribution whether or not a trend is included or not
C) is a two-sided test
D) is an extension of the Dickey-Fuller test when the underlying model is AR(p) rather than AR(1)

Continued...

19. An autoregression is a regression _____.
A) of a dependent variable on lags of regressors
B) that allows for the errors to be correlated
C) model that relates a time series variable to its past values
D) to predict sales in a certain industry
20. The Fixed Effects regression model _____.
A) has n different intercepts
B) the slope coefficients are allowed to differ across entities, but the intercept is "fixed" (remains unchanged)
C) has "fixed" (repairs) the effect of heteroskedasticity
D) in a log-log model may include logs of the binary variables, which control for the fixed effects
21. Time Fixed Effects regression is useful in dealing with omitted variables _____.
A) even if you only have a cross-section of data available
B) if these omitted variables are constant across entities but vary over time
C) when there are more than 100 observations
D) if these omitted variables are constant across entities but not over time
22. Analysing the behaviour of unemployment rates across U.S. states in March of 2006 is an example of using _____.
A) time series data
B) panel data
C) cross-sectional data
D) experimental data
23. Panel data _____.
A) is also called longitudinal data
B) is the same as time series data
C) studies a group of people at a point in time
D) typically uses control and treatment groups
24. To provide quantitative answers to policy questions _____.
A) it is typically sufficient to use common sense
B) you should interview the policy makers involved
C) you should examine empirical evidence
D) is typically impossible since policy questions are not quantifiable
25. The main advantage of using panel data over cross-sectional data is that it _____.
A) gives you more observations
B) allows you to analyse behaviour across time but not across entities
C) allows you to control for some types of omitted variables without actually observing them
D) allows you to look up critical values in the standard normal distribution

Continued...

SECTION B (STRUCTURAL QUESTIONS)**QUESTION 1 (25 marks)**

(a) Interpret the consequences of the following econometric problems:

- (i) Imperfect multicollinearity (3 marks)
- (ii) Serial correlation (3 marks)
- (iii) Heteroscedasticity (3 marks)

(b) Consider the following money supply model in Malaysia

$$M2_t = \beta_0 + \beta_1 GDP_t + \beta_2 CPI_t + \beta_3 INT_t + u_t \quad \dots \quad (1.1)$$

where $M2$ = money supply (in RM)
 GDP = gross domestic product (in RM)
 CPI = consumer price index
 INT = interest rate (in percentage)
 u = error term

The following outputs (Tables 1.1 and 1.2) are obtained from EViews:

Table 1.1 Estimation Output

Dependent Variable: M2

Method: Ordinary Least Square (OLS)

Included observations: 36

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP	1.523689	0.056492	26.97193	0.0000
CPI	-2030.989	302.2391	-6.719808	0.0000
INT	26.48578	1376.122	0.019247	0.9848
C	52540.41	13024.88	4.033849	0.0003
R-squared	0.993898	Mean dependent var	145484.6	
Adjusted R-squared	0.993326	S.D. dependent var	166864.1	
S.E. of regression	13632.18	Akaike info criterion	21.98269	
Sum squared resid	5.95E+09	Schwarz criterion	22.15864	
Log likelihood	-391.6885	Hannan-Quinn criter.	22.04410	
F-statistic	1737.335	Durbin-Watson stat	0.959108	
Prob(F-statistic)	0.000000			

Continued...

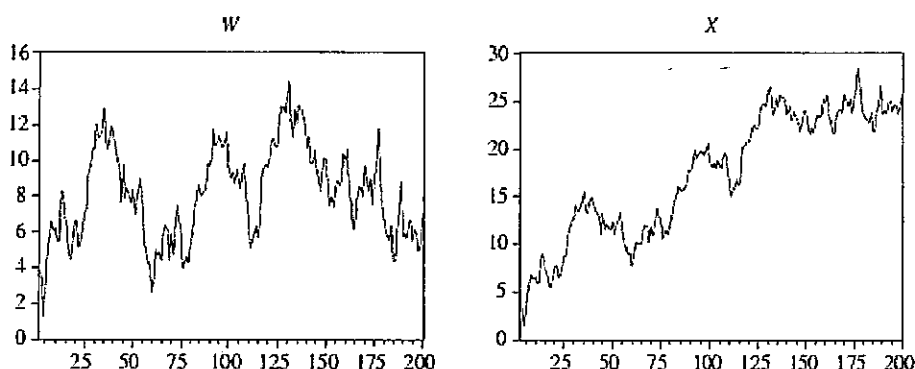
Table 1.2 Correlation Matrix

	M2	GDP	CPI	INT
M2	1.000000	0.991721	0.896999	-0.456038
GDP	0.991721	1.000000	0.939649	-0.421299
CPI	0.896999	0.939649	1.000000	-0.266809
INT	-0.456038	-0.421299	-0.266809	1.000000

- (i) From the output, demonstrate how you detect the econometric problems. (4 marks)
- (ii) Given your answer in b (i), assess the consequential impacts on the properties of the least-squares estimators and the estimated standard errors if there is an econometric problem detected in the model. (4 marks)
- (iii) Show **TWO (2)** remedies that you may use to solve the problem that arises from the estimation of the above model. (8 marks)

QUESTION 2 (25 marks)

- (a) Figure 2.1 shows plots of two time series.

Figure 2.1 Time series plot

Since W appears to be fluctuating around a nonzero mean, a Dickey–Fuller test (with constant but no trend) was performed on this variable. Since X appears to be fluctuating around a trend, a Dickey–Fuller test (with constant and trend) was performed for this variable.

Continued...

The results are shown below.

$$\begin{array}{rcl} \widehat{\Delta W}_t & = & 0.757 - 0.091W_{t-1} \\ \tau & & (2.154) \quad (-3.178) \end{array} \quad \dots \quad (2.1)$$

$$\begin{array}{rcl} \widehat{\Delta X}_t & = & 0.782 + 0.171t - 0.092X_{t-1} \\ \tau & & (1.867) \quad (1.554) \quad (-3.099) \end{array} \quad \dots \quad (2.2)$$

The 5 percent τ critical values are -3.41 (constant with trend), and -2.86 (constant but no trend).

- (i) Prepare the null and alternative hypotheses for Dickey-Fuller test. (2 marks)
 - (ii) Interpret on which series is/ are stationary, and which is/ are nonstationary. (5 marks)
- (b) Utilising quarterly data on disposable income (INCOME) and consumption (CONSUME) in Malaysia from 2007Q1 to 2021Q4, you obtain the following results:

$$\begin{array}{rcl} \widehat{\text{CONSUME}}_t & = & -9084 + 0.9884\text{INCOME} \\ t & & (-2.154) \quad (9.191) \end{array} \quad \dots \quad (2.3)$$

$$\begin{array}{rcl} \widehat{\Delta e}_t & = & -0.316\widehat{e}_{t-1} \\ \tau & & (-3.909) \end{array} \quad \dots \quad (2.4)$$

Based on the above estimation, given the 5 percent τ critical value of -3.37 , interpret if the CONSUME is cointegrated with, or spuriously related to INCOME if

- (i) CONSUME is I(2) and INCOME is I(1) variables (4 marks)
- (ii) CONSUME and INCOME are I(1) variables (4 marks)

Continued...

- (c) Based on the Michigan Income Dynamics Study, Hausman attempted to estimate a wage, or earnings, model using a sample of 629 high school graduates, who were followed for a period of 6 years, thus giving in all 3774 observations. The dependent variable in this study was logarithm of wage, and the explanatory variables were age (divided into several age groups), unemployment in the previous year, poor health in the previous year, self-employment, region of residence (South = 1; 0 otherwise), area of residence (rural = 1; 0 otherwise). Hausman used both fixed effects model (FEM) and random effects model (REM). The results are given in **Table 2.1** (standard errors in parentheses):

Table 2.1 Wage Equations (dependent Variable: Log Wage*)

Variable	Fixed effects		Random effects	
	Estimated Coefficient	Standard error	Estimated Coefficient	Standard error
Age 1 (20-34)	0.0557	(0.0042)	0.0393	(0.0033)
Age 2 (35-44)	0.0351	(0.0051)	0.0092	(0.0036)
Age 3 (45-54)	0.0209	(0.0055)	-0.0007	(0.0042)
Age 4 (55-64)	0.0209	(0.0078)	-0.0097	(0.0060)
Age 5 (65 and above)	-0.0171	(0.0155)	-0.0423	(0.0121)
Unemployed previous year	-0.0042	(0.0153)	-0.0277	(0.0151)
Poor health previous year	-0.0204	(0.0221)	-0.0250	(0.0215)
Self-employment	-0.2190	(0.0297)	-0.2670	(0.0263)
South	-0.1599	(0.0656)	-0.0324	(0.0333)
Rural	-0.0101	(0.0317)	-0.1215	(0.0237)
Constant	-	-	0.8499	(0.0433)
s^2	0.0567		0.0694	
Degree of freedom	3,135		3,763	

Note: *3774 observations.

Reproduced from Cheng Hsiao, *Analysis of Panel Data*, Cambridge University Press, 1986, p.42. Original source: J.A. Hausman, "Specification Tests in Econometrics," *Econometrica*, vol 46, 1978, pp. 1251-1271.

- (i) On the basis of the data given in the **Table 2.1**, which model, if any, would you choose? Interpret your selection. (4 marks)
- (ii) Demonstrate the hypothesis and steps involved in choosing fixed effects or random effects model using the Hausman test. (6 marks)

End of Page

Formulas

$$T \text{ Statistics} = \frac{\hat{\beta} - \beta}{SE(\hat{\beta})}$$

$$F \text{ Statistics} = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$R^2 = \frac{ESS}{TSS}$$

$$\text{Adjusted } R^2 = 1 - (1 - R^2) \left(\frac{n-1}{n-k} \right)$$

Total Sum of Squares (TSS) = Explained Sum of Squares (ESS) + Residual Sum of Squares (RSS)

White's general heteroscedasticity test statistics, $\chi^2_{df=k} = n \times R^2$

$$\rho \approx 1 - (d/2)$$

Statistical Tables

Appendix 1: t-Table

two tails	0.2	0.1	0.05	0.02	0.01
One tail	0.1	0.05	0.025	0.01	0.005
df					
10	1.37	1.81	2.23	2.76	3.17
20	1.33	1.72	2.09	2.53	2.84
30	1.31	1.70	2.04	2.46	2.75
40	1.30	1.68	2.02	2.42	2.70
50	1.30	1.68	2.01	2.40	2.68
60	1.30	1.67	2.00	2.39	2.66
75	1.29	1.67	1.99	2.38	2.64
100	1.29	1.66	1.98	2.36	2.63
120	1.29	1.66	1.98	2.36	2.62

Appendix 2: F-table ($\alpha=0.05$)

df2\df1	1	2	3	4	5	6
10	4.96	4.10	3.71	3.48	3.33	3.22
20	4.35	3.49	3.10	2.87	2.71	2.60
30	4.17	3.32	2.92	2.69	2.53	2.42
40	4.08	3.23	2.84	2.61	2.45	2.34
50	4.03	3.18	2.79	2.56	2.40	2.29
60	4.00	3.15	2.76	2.53	2.37	2.25
70	3.98	3.13	2.74	2.50	2.35	2.23
80	3.96	3.11	2.72	2.49	2.33	2.21
100	3.94	3.09	2.7	2.46	2.31	2.19
200	3.89	3.04	2.65	2.42	2.26	2.14
500	3.86	3.01	2.62	2.39	2.23	2.12
1000	3.85	3.00	2.61	2.38	2.22	2.11

Appendix 3: Chi-square table ($\alpha=0.05$)

df	$\alpha = 0.05$
1	3.84
2	5.99
3	7.82
4	9.49
5	11.07
10	18.31
20	31.41
50	67.51
100	124.34

Appendix 4: Durbin Watson d Table

Critical Values for the Durbin-Watson Statistic (d)										
Level of Significance $\alpha = 0.05$										
<i>n</i>	<i>k</i> = 1		<i>k</i> = 2		<i>k</i> = 3		<i>k</i> = 4		<i>k</i> = 5	
	d _L	d _U	d _L	d _U	d _L	d _U	d _L	d _U	d _L	d _U
10	0.88	1.32	0.70	1.64	0.53	2.02	0.38	2.41	0.24	2.82
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
25	1.29	1.41	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77

where *n* = number of observations and *k* = number of independent variables, excluding the intercept.

Appendix 5: 1% and 5% Critical Dickey Fuller ($\tau=\tau$) for Unit Root Tests

Sample size	τ_{nc}		τ_c		τ_{ct}	
	1%	5%	1%	5%	1%	5%
25	-2.66	-1.95	-3.75	-3.00	-4.38	-3.60
50	-2.62	-1.95	-3.58	-2.93	-4.15	-3.50
100	-2.60	-1.95	-3.51	-2.89	-4.04	-3.45
250	-2.58	-1.95	-3.46	-2.88	-3.99	-3.43
500	-2.58	-1.95	-3.44	-2.87	-3.98	-3.42
∞	-2.58	-1.95	-3.43	-2.86	-3.96	-3.41

Note: Subscripts nc, c and ct denote, respectively, that there is no constant, a constant, and a constant and trend term in the regression.

